PTAS for MAP Assignment on Pairwise Markov Random Fields in Planar Graphs

Eli Fox-Epstein, Roie Levin, and David Meierfrankenfeld

Department of Computer Science, Brown University, Providence RI.
{ef,rl46,nfelddav}@cs.brown.edu

Abstract. We present a PTAS for computing the maximum a posteriori assignment on Pairwise Markov Random Fields with non-negative weights in planar graphs. This algorithm is practical and not far behind state-of-the-art techniques in image processing. MAP on Pairwise Markov Random Fields with (possibly) negative weights cannot be approximated unless P = NP, even on planar graphs. We also show via reduction that this yields a PTAS for one scoring function of Correlation Clustering in planar graphs.

1 Introduction

Pairwise Markov Random Fields (MRFs) model distributions in a variety of applications and arise in fields as diverse as statistical physics, computer vision, coding theory, computational biology, machine learning, and combinatorial optimization. Solving associated optimization problems is critical in practice and also of high theoretical importance.

We briefly review the statistical view on MRFs before focusing on the combinatorial problem. A pairwise MRF is a set of n random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ over label set $\{1, \ldots, L\}$, a graph $G = (\boldsymbol{X}, E)$, where

$$\Pr[\boldsymbol{X} = \boldsymbol{x}] = \frac{1}{Z} \exp\left(\sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \psi_{ij}(x_i, x_j)\right),$$

where ϕ_i and ψ_{ij} are arbitrary functions and Z is a normalizing constant. Intuitively, $\phi_i(x_i)$ can be regarded as vertex i's preference for label x_i and $\psi_{ij}(x_i, x_j)$ as the compatibility between labels x_i and x_j on the endpoints of edge ij. We are interested in finding a maximum a posteriori (MAP) assignment $\boldsymbol{x}^*$, i.e. $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} \Pr[\boldsymbol{X} = \boldsymbol{x}]$. Finding the MAP label assignment corresponds to this optimization problem:

Pairwise MAP MRF

Instance:
- graph $G = (V, E)$,
- label set $\mathcal{L} = \{1, \dots, L\}$,
- singleton functions $\phi_i(\cdot) : \mathcal{L} \to \mathbb{R} \quad \forall\, i \in V$,
- pairwise functions $\psi_{ij}(\cdot, \cdot) : \mathcal{L} \times \mathcal{L} \to \mathbb{R} \quad \forall\, (i, j) \in E$.

Solution: for each $v \in V$, label assignment $x_v \in \mathcal{L}$

Maximize:

$$H(\boldsymbol{x}) = \sum_{v \in V} \phi_i(x_v) + \sum_{(u,v) \in E} \psi_{uv}(x_u, x_v),$$

Throughout the paper, we will assume G to be connected; combining the solutions on each component handles the case of a disconnected graph.

Pairwise MAP MRF has been considered in many domains, however, in general graphs we will show:

Theorem 1. *There is an $\alpha > 0$ such that, unless P = NP, there is no polynomial-time α-approximation algorithm for* Pairwise MAP MRF*, even for nonnegative ϕ and ψ.*

In light of this, we focus on planar graphs as many real-world instances, such as those from computer vision, are planar or nearly planar. It turns out that Pairwise MAP MRF is still NP-hard on planar graphs [2]. However, restricting our attention to planar graphs allows for much better approximation algorithms.

We will additionally require ϕ_i and ψ_{ij} to be nonnegative. By setting

$$\phi'_i(x) = \phi_i(x) - \min_{a \in \mathcal{L}}(\phi_i(a)) \qquad \forall\, i \in V, x \in \mathcal{L} \qquad \text{and}$$

$$\psi'_{ij}(x, y) = \psi_{ij}(x, y) - \min_{a,b \in \mathcal{L}}(\psi_{ij}(a, b)) \qquad \forall\, (i, j) \in E,\ \forall\, x, y \in \mathcal{L},$$

we can transform an instance with general weights into an instance with nonnegative weights with the same optimal assignment. However, this changes the value of the objective function, and thus also the approximation ratio. This restriction is necessary, as with general weights Pairwise MAP MRF is impossible to approximate unless P = NP. In particular:

Theorem 2. *The existence of an algorithm approximating* Pairwise MAP MRF *on planar graphs with maximum degree 4 and nonpositive ϕ_i and ψ_{ij} to any multiplicative factor implies P = NP.*

In many applications, MRF is used to minimize an energy function. Notice that this is equivalent to maximizing the negative energy function. Thus Theorem 2 implies minimization is inapproximable to any multiplicative factor, even if the energy function is nonnegative.

A *polynomial-time approximation scheme* (PTAS) is an algorithm that, given an instance of a maximization (minimization) problem and a precision parameter $0 < \varepsilon < 1$, returns a $(1 - \varepsilon)$-approximate ($(1 + \varepsilon)$-approximate, resp.)

solution in time polynomial in the size of the instance (with a possible exponential dependence on $1/\varepsilon$). An *efficient* PTAS (EPTAS) is one with runtime of the form $O(f(\varepsilon)\,\mathrm{poly}(n))$, where n is the size of the instance and f is a computable function.

Our main result is:

Theorem 3. *There is a PTAS for* PAIRWISE MAP MRF *in planar graphs when all ϕ and ψ are nonnegative functions.*

We also consider the closely related CORRELATION CLUSTERING problem. In this, one is given a graph and tasked with partitioning the vertices into an arbitrary number of clusters. The edges have associated rewards and preferences as to whether their endpoints should or should not be in the same cluster; the objective function is the sum of the weights of the edges whose preferences are satisfied. CORRELATION CLUSTERING is sometimes expressed with a penalty for unsatisfied edges in addition to, or instead of, a reward for satisfied edges. These formulations all have the same optimal solution, but as in PAIRWISE MAP MRF, the value of the objective function changes, and thus approximation results may differ as well.

Formally, the version we will address is:

CORRELATION CLUSTERING

Instance:
- graph $G = (V, E)$,
- edge preferences $p : E \to \{0, 1\}$,
- edge reward function $w : E \to \mathbb{R}_{\geq 0}$.

Solution: a partition of the vertices into clusters.
Maximize:

$$\sum_{(u,v)\in E} w(u,v)\left[(1-p(u,v))C(u,v) + p(u,v)(1-C(u,v))\right]$$

where $C(u,v)$ is 1 if u and v belong to the same cluster and 0 otherwise.

Via a simple reduction to PAIRWISE MAP MRF:

Corollary 1. *There is an EPTAS for* CORRELATION CLUSTERING *in planar graphs.*

1.1 Outline

In Section 2, we review past work on PAIRWISE MAP MRF. Next, in Section 3, we give an exact algorithm for graphs of bounded branchwidth. Then, Section 4 proves Theorem 3. In the interest of space, proofs of Corollary 1 and Theorems 1 and 2 are in the appendix. We demonstrate some promising experimental results in Section 5 with applications to computer vision. Finally, we offer discussion in Section 6.

2 Prior Work

Markov Random Fields originated in statistical physics as a generalization of the Ising Model [15]. There are numerous techniques to solve PAIRWISE MAP MRF, both in general and on specific instances; some are outlined here.

An MRF is *binary* if there are exactly two labels and *submodular* if for all $u, v \in V$, for all $i, j \in \{1, \ldots, L\}$, $\psi_{u,v}(i,i) + \psi_{u,v}(j,j) \geq \psi_{u,v}(i,j) + \psi_{u,v}(j,i)$.

If an MRF is both binary and submodular, PAIRWISE MAP MRF can be solved exactly in polynomial time by reduction to MIN-CUT [5]. If the graph is also planar, the running time can be improved to $O(n \log(n))$ [20].

For MRF in graphs which have bounded degree and an excluded minor (which includes all bounded degree planar graphs) Jung and Shah use techniques similar to ours to find a PTAS with running time doubly exponential in $1/\varepsilon$ [12]. For the alternate formulation of CORRELATION CLUSTERING which seeks to minimize penalties for unsatisfied edges, Klein et al. demonstrate a non-efficient PTAS in planar graphs [17].

When ψ_{ij} are defined by a metric on the labels, the problem is referred to as METRIC LABELING; [18] provides a $O(\log L \log\log L)$-approximation algorithm for the problem.

The GENERALIZED POTTS MODEL, from statistical mechanics, is a restriction of MRF that reduces to the classic MULTIWAY CUT problem; [6] uses local search to approximate this model. MULTIWAY CUT is a special case of METRIC LABELING, where some vertices are forced to have particular labels. In planar graphs, there is a PTAS for the problem [3]. In general, there are constant-factor approximations [8].

0-EXTENSION is a generalization of MULTIWAY CUT in which the cost of the edge depends on the specific terminals associated with the edge's endpoints, not just whether the terminals are the same. In general graphs, this problem is $O(\log L / \log\log L)$-approximable [14,7,11] and can be approximated to a constant-factor in planar graphs.

Various heuristics exist to approximate MAP on planar graphs and are used extensively in computer vision for applications such as:

- Stereo vision: given two photographs taken side-by-side, estimate the depths of each pixel.
- Object segmentation: find the boundaries of objects in photographs.
- De-noising: remove grainy noise from an image.
- Photomontage: combine several images into one.

Two standard benchmarks for these problems are OpenGM [13] and the Middlebury stereo dataset [19]. For a detailed treatment of MRF as applied to computer vision, see, e.g., [24].

Many problems, including PAIRWISE MAP MRF and more traditional optimization problems such as TSP, STEINER TREE, VERTEX COVER, GRAPH COLORING, CLIQUE, HAMILTONIAN PATH, and FEEDBACK VERTEX SET can be solved exactly in polynomial time on graphs of bounded *branchwidth*. Branchwidth, like treewidth, pathwidth, bandwidth, outerplanarity, or cliquewidth, is a

measure of the "simplicity" of a graph. These measures are amenable to dynamic programming and have been of great importance when designing approximation schemes on planar graphs [1,10,16,4].

Our algorithm draws inspiration from Baker's technique [1], a powerful framework for building PTASes in planar graphs. In a nutshell, Baker guesses a way to decompose a graph into a number of smaller graphs of bounded outerplanarity. These smaller graphs are each solved optimally and independently, and then combining the solutions incurs at most $\varepsilon\,\mathrm{OPT}$ error. This technique was originally applied to Independent Set but can be used for a number of problems, such as Vertex Cover, Edge-Disjoint Triangles, and Dominating Set [1].

Recently, Wang posted a manuscript on arXiv claiming a PTAS for Pairwise MAP MRF on planar graphs, among other results [23]. We remark that our main result, Theorem 3, was discovered independently. Theorem 2 draws inspiration from and strengthens a hardness proof of Wang. Unfortunately, there appears to be a bug in an vital lemma in [23]. We discuss this in Appendix B.

3 Pairwise MAP MRF in Bounded Branchwidth Graphs

A branch decomposition of a graph $G = (V, E)$ is an unrooted binary tree T whose leaves are the edges E of G. Deleting an edge of T generates two subgraphs of G, each induced by the edges in one component of T. Some vertices are contained in both subgraphs. The maximum number of these overlapping vertices for any such pair of subgraphs is the *width* of the decomposition. The minimum width of any branch decomposition of G is its *branchwidth*.

Our PTAS is an application of Baker's technique [1], and works by breaking up the problem into bounded branchwidth subproblems, each of which can be solved exactly in polynomial time.

Theorem 4. *Given an* Pairwise MAP MRF *instance* $(G = (V, E), L, \phi, \psi)$ *and a branch decomposition* T *of width* k*, an optimal solution can be found in time* $O(|E|kL^{2k})$.

Proof. We use dynamic programming. T will guide the dynamic program, and thus we want a root with two children. To that end, we choose an arbitrary edge of T and subdivide it with a new vertex r that we designate the root. Now T is a rooted binary tree but maintains the other properties of a branch decomposition.

With each tree vertex $v \in T$, let $G(v)$ be the subgraph of G induced by the edges of G which are descendants of v. Observe that $G(r) = G$. Denote by $\delta(G(v))$ the vertices of $G(v)$ which are incident to edges not in $G(v)$. Note that $|\delta(G(v))| \leq k$ for all $v \in T$.

For each vertex $v \in T$, we will compute the assignment to the vertices $V(G(v)) - \delta(G(v))$ for each possible assignment to the vertices $\delta(G(v))$ which maximizes the score of the MRF on $G(v)$. This is done bottom-up, so that for all non-leaf vertices of T, assignments for both of their children are computed first.

If v is a leaf, $G(v)$ is a single edge with its endpoints. Thus either $V(G(v)) - \delta(G(v))$ is empty and finding the optimal assignment is trivial; or $V(G(v)) - \delta(G(v))$ is a single endpoint, and all possible label assignments can be tested. In both cases, it takes $O(L^2)$ time to test for all possible boundary assignments what the best assignment to $V(G(v)) - \delta(G(v))$ is.

If v is not a leaf, it has two children u_1, u_2. Let $U = \delta(G(u_1)) \cup \delta(G(u_2))$ and $I = \delta(G(u_1)) \cap \delta(G(u_2))$. Notice $\delta(G(v)) \subseteq U$. For each label assignment to the vertices of $\delta(G(v))$, the best assignment to $V(G(v)) - \delta(G(v))$ is the union of best assignments to $V(G(u_1)) - \delta(G(u_1))$ and $V(G(u_1)) - \delta(G(u_1))$ for some assignment to $I - \delta(G(v))$, and its value is the sum of the values of those assignments minus the values of ϕ on I. As those assignments and values have already been computed, finding the optimal ones can be done in time $O(|I|L^{|U|})$. Since $|I| \leq k$ and $|U| \leq 2k$, computing all the assignments and values at vertex v takes time $O(k^{2k})$.

$\delta(G(r))$ is empty, so the unique assignment and value computed at r are the exact optimal solution to the Pairwise MAP MRF instance. The rooted branch decomposition has $2|E|-1$ vertices, thus the running time is $O(|E|kL^{2k})$. □

We summarize the algorithm:

1. Choose an arbitrary edge e of T, and subdivide it with a new root vertex r.
2. With each vertex v of T associate the subgraph $G(v)$ of G induced by the edges of G which are descendants of v (with respect to the root r).
3. Consider each vertex v of T from leaf to root:
   (a) If v is a leaf, for each possible label assignment to the vertices of $\delta(G(v))$, by brute force, compute the best assignment to $V(G(v)) \setminus \delta(G(v))$.
   (b) Otherwise, for each possible label assignment to the vertices of $\delta(G(v))$, combine the values and assignments of v's two children to determine the best assignment to $V(G(v)) \setminus \delta(G(v))$.
4. Return the best assignment for $G(r) = G$.

4 PTAS for Pairwise MAP MRF on Planar Graphs

We now give the PTAS for our main result. As input, we are given an instance of Pairwise MAP MRF $\langle G = (V, E), L, \phi, \psi \rangle$ where G is a planar graph, and a desired error parameter $0 < \varepsilon < 1$, with $k = \frac{1}{\varepsilon}$.

Fix some vertex r. We say an edge has r-level d if one of its endpoints is hop-distance $d-1$ from r and the other is hop-distance d. Let G_j be the graph resulting in deleting all edges with r-levels congruent to $j \pmod k$.

The algorithm is:

1. Choose a vertex r arbitrarily.
2. Let $k = \frac{1}{\varepsilon}$.
3. For each $j \in \{0, \ldots, k-1\}$:
   (a) Compute G_j.
   (b) Find an approximate branch decomposition T of each component of G_j using the algorithm in [21].
   (c) Apply Theorem 4 to each component of G_j and combine the resulting best label assignments into $\boldsymbol{x}_j$.
   (d) Compute the value h_j of the objective function on G from $\boldsymbol{x}_j$.
4. Return the assignment corresponding to the largest h_j.

With this, we are ready to prove our main result.

Proof (of Theorem 3). First, we tackle the runtime. For each j, it takes linear time to construct G_j by building a breadth first search tree from r.

By construction, there exists in each component of G_j a path of length at most k from each vertex to a vertex on the face containing r. An algorithm by Tamaki [21] allows us to construct a branch decomposition of width at most $2k$ on a graph with this property in time $O(m_i 2^{2k})$, where m_i is the number of edges in the component.

Then, solving these optimally using 4 and combining takes time $O(|E|kL^{4k})$. As we try k different choices of j, the total running time is $O(|E|k^2L^{4k})$. This is linear in the size of the graph, as k is a function of ε. However, as L is part of the input, this is not an efficient PTAS.

Now, we demonstrate correctness. Let $\boldsymbol{x}^*$ be an optimal label assignment. By construction, $\boldsymbol{x}_j$ is the optimal assignment on G_j. Let H_j be the objective function restricted to G_j. Since $\boldsymbol{x}_j$ consists of optimal solutions of each component of G_j, $H_j(\boldsymbol{x}_j) \geq H_j(\boldsymbol{x}^*)$.

Let $d_j = H(\boldsymbol{x}^*) - H_j(\boldsymbol{x}^*)$. So we have $H(\boldsymbol{x}_j) \geq H(\boldsymbol{x}^*) - d_j$. Summing over all choices of j,

$$\sum_{j=0}^{k-1} H(\boldsymbol{x}_j) \geq \sum_{j=0}^{k-1} H(\boldsymbol{x}^*) - d_j(\boldsymbol{x}^*).$$

Each edge in G is missing from at most one G_j, so $\sum_{j=0}^{k-1} d_j \leq H(\boldsymbol{x}^*)$. Thus,

$$\sum_{j=0}^{k-1} H(\boldsymbol{x}_j) \geq kH(\boldsymbol{x}^*) - H(\boldsymbol{x}^*) = k(1-1/k)H(\boldsymbol{x}^*) = k(1-\varepsilon)H(\boldsymbol{x}^*).$$

Consequently, there exists some j where $H(\boldsymbol{x}_j) \geq (1-\varepsilon)H(\boldsymbol{x}^*)$. □

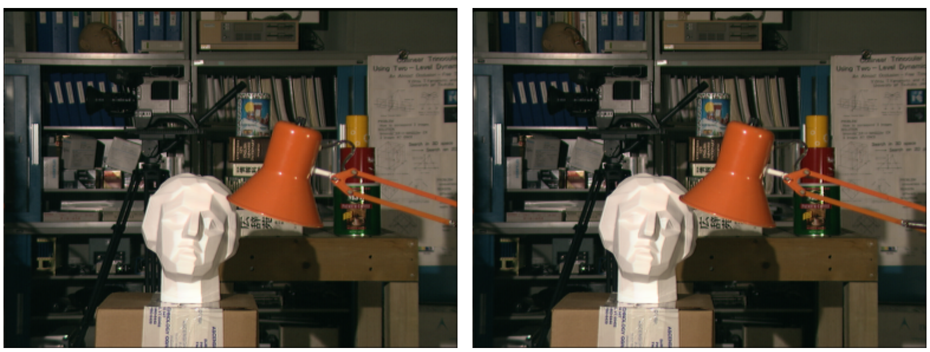
5 Experiments

The approximation scheme has relatively small constants, which suggested that it might be feasible to use in practice. We implemented a version of this PTAS

in `C++11` for tasks that arise in computer vision. For simplicity, we restricted our implementation to grid graphs, as is common in image processing. Optimal branch decompositions are particularly easy to find in this domain.

5.1 Stereo Matching

Given two images representing a left camera angle and a right camera angle and a number L of relative depth labels, we wish to assign a label in $\{1, \ldots, L\}$ to each pixel in the, say, left image. In the computer vision community, these are often visualized as *disparity maps*, or grayscale images of the relative depths; see e.g. Figure 2. We use the 16 label *tsukuba* example from the Middlebury stereo benchmark [19] for illustration here:



(a) Left Image (b) Right image

Fig. 1: *Tsukuba* images from the Middlebury stereo benchmark.

We used the following model as input to our algorithm. The graph $G = (V, E)$ is the planar grid graph where each vertex represents a pixel. We define functions

$$\phi_u(i) = \beta - \|u - u^{(-i)}\|_2^2 \qquad \forall\, u \in V$$

$$\psi_{u,v}(i,j) = \begin{cases} 0 & \text{if } i = j \\ \beta - \|u - v\|_2^2 & \text{if } i \neq j \end{cases} \qquad \forall\, (u,v) \in E$$

where u is a pixel in the left image, $u^{(-i)}$ is the pixel that is i columns to the left of the pixel corresponding to u in the right image, $\| \cdot \|_2^2$ is square 2-norm in CIELUV color space, and β is a constant sufficiently large to ensure that all outputs of the functions are positive.

In addition to our basic algorithm, we also incorporate a few very simple vision-specific heuristics to refine our results. Initializing boundary pixels to the values from the previous (either left or right) connected component yields more visually continuous results. Since the analysis of the approximation holds for any

value of the boundary pixels, in particular it holds for these values. Thus the approximation guarantee is preserved at this step. However, this results in some visual artifacts (see Figure 2).

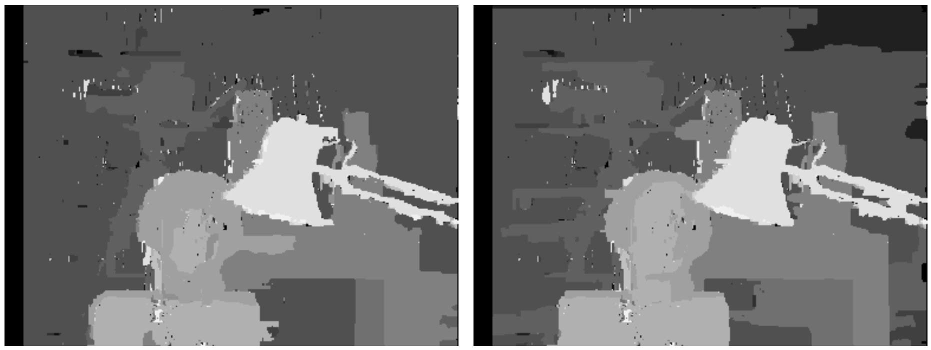


Fig. 2: Visual artifacts after one heuristic.

To remedy this, we run the algorithm twice (intuitively, left-to-right and then right-to-left) and combine the solutions in an approximation-preserving way.

Finally, a tiny amount of smoothing is done to remove remaining noise; this does not guarantee the approximation but leads to more visually-pleasing results.

We use the evaluation tools provided on the Middlebury stereo website: 5.07% of all pixels are mislabeled including 3.02% of non-occluded regions and 11.5% of regions near depth discontinuities. Furthermore, as seen in Figure 4d, a large fraction of mislabeled pixels are concentrated in the bottom right; we believe that discrepancies between the MRF model and the ground truth explain this.

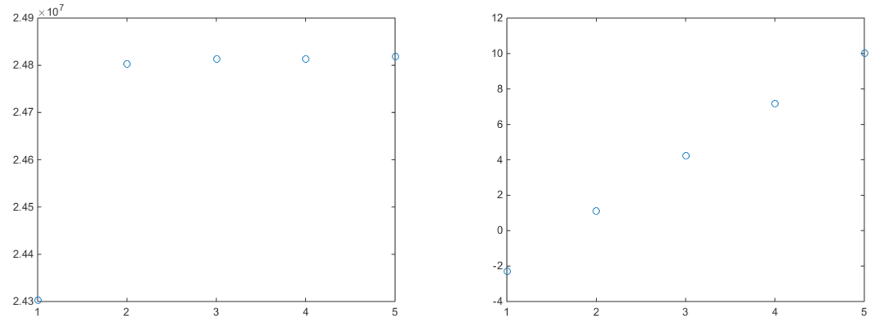
State-of-the-art algorithms mislabel a little more than 1% of pixels including typically over 4% of regions near depth discontinuities. Many of the published algorithms on the Middlebury benchmark mislabel significantly more than 5.07% of all pixels, and the best algorithms involve optimizing dozens of hyperparameters and are highly specialized to their applications.

We found that our generic PTAS required only a few basic heuristics to perform quite well, suggesting that with a few more heuristics, it could be very competitive.

5.2 Observed ε dependencies

Experiments support the theoretical dependencies on ε. Figures 3a and 3b show the score and log running time, respectively, of our algorithm on the *tsukuba* image as a function of $1/\varepsilon$, using 14 labels and the learned parameters. The score changes remarkably little, considering the improvement in the theoretical bound.

The running time matches the theory very closely. The observed ratios of running time as ε increases from 2 to 5 are 23, 18.45, and 17.8; the theoretical



(a) Score (in arbitrary units) as a function of $1/\varepsilon$.

(b) Log of running time (in seconds) as a function of $1/\varepsilon$. Experiments were run on a mid-range 2014 laptop.

Fig. 3

run time is proportional to $1/\varepsilon \cdot L^{1/\varepsilon}$ which would predict ratios of 21, 18.67, and 17.5.

5.3 OpenGM benchmark

We used the OpenGM 2.3.3 [13] library to benchmark the actual energy minimization performance of our algorithm compared to other existing methods. Our algorithm was run with $\varepsilon = 1/3$.

On the Inpainting benchmark, our algorithm achieves a score of 461.82, which is about 1.6% away from the best algorithm's and better than half of the competing algorithms. On the Object Segmentation benchmark, we perform a bit worse; our score is about 64% away from the best and worse than most of the competition.

6 Discussion & Conclusions

Our algorithm gives the first known PTAS for maximum a posteriori assignment on Pairwise MAP MRF, and the first EPTAS for this variant of Correlation Clustering in planar graphs. Combined with our hardness results, much of the complexity of Pairwise MAP MRF on planar graphs is now settled. While the algorithm is not directly competitive with the state of the art for computer vision tasks, it is sufficiently close to those algorithms to suggest applications in improving them, as well as in other applications which lack specialized algorithms.

One can readily extend the given PTAS to more general classes of graphs, or (non-pairwise) MRFs in planar graphs with bounded factor degree.

Compelling future research directions include studying Pairwise MAP MRF with with negative functions and two labels (but not necessarily submodular), and with more than two labels but submodular functions.

References

1. B. S. Baker. Approximation algorithms for NP-complete problems on planar graphs. In *Foundations of Computer Science, 1983., 24th Annual Symposium on*, pages 265–273, Nov 1983.
2. F. Barahona. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General*, 15(10):32–41, 1982.
3. M. Bateni, M.i Hajiaghayi, P. N. Klein, and C. Mathieu. A polynomial-time approximation scheme for planar multiway cut. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 639–655. SIAM, 2012.
4. G. Borradaile, P. N. Klein, and C. Mathieu. Steiner tree in planar graphs: An o(nlogn) approximation scheme with singly-exponential dependence on epsilon. In *Algorithms and Data Structures*, volume 4619 of *Lecture Notes in Computer Science*, pages 275–286. Springer Berlin Heidelberg, 2007.
5. Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(9):1124–1137, Sept 2004.
6. Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *Proceedings of the 1998 Conference on Computer Vision and Pattern Recognition*, pages 648–655, 1998.
7. G. Calinescu, H. Karloff, and Y. Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the 12th Symposium on Discrete Algorithms*, pages 8–16, 2001.
8. Elias Dahlhaus, David S. Johnson, Christos H. Papadimitriou, Paul D. Seymour, and Mihalis Yannakakis. The complexity of multiway cuts (extended abstract). In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 241–251, 1992.
9. D. P. Dailey. Uniqueness of colorability and colorability of planar 4-regular graphs are np-complete. *Discrete Mathematics*, 30(3):289 – 293, 1980.
10. E. D. Demaine and M. Hajiaghayi. Bidimensionality: New connections between fpt algorithms and ptass. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 590–601, Philadelphia, PA, USA, 2005. SIAM.
11. J. Fakcharoenphol, C. Harrelson, S. Rao, and K. Talwar. An improved approximation algorithm for the 0-extension problem. In *Proceedings of the 13th Symposium on Discrete Algorithms*, pages 257–265, 2003.
12. K. Jung and D. Shah. Local algorithms for approximate inference in minor-excluded graphs. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 729–736. Curran Associates, Inc., 2008.
13. J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, T. Kröger, J. Lellmann, N. Komodakis, B. Savchynskyy, and C. Rother. A comparative study of modern inference techniques for structured discrete energy minimization problems. *CoRR*, abs/1404.0533, 2014.
14. A. Karzanov. Minimum 0-extensions of graph metrics. *European Journal of Combinatorics*, 19(1):71–101, 1998.
15. R. Kindermann and J. L. Snell. Markov Random Fields and Their Applications. In *Contemporary Mathematics*, June 1980.

16. P. N. Klein. A linear-time approximation scheme for planar weighted tsp. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 647–656, Oct 2005.
17. P. N. Klein, C. Mathieu, and H. Zhou. Correlation clustering and two-edge-connected augmentation for planar graphs. In *Proceedings of the 32nd Symposium on Theoretical Aspects of Computer Science*, 2015. to appear.
18. J. M. Kleinberg and E. Tardos. Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. *Journal of the ACM*, 49(5):616–639, 2002.
19. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1):742, May 2002.
20. F. R. Schmidt, E. Toppe, and D. Cremers. Efficient planar graph cuts with applications in computer vision. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 351–356, June 2009.
21. H. Tamaki. A linear time heuristic for the branch-decomposition of planar graphs. In Giuseppe Di Battista and Uri Zwick, editors, *Algorithms - ESA 2003*, volume 2832 of *Lecture Notes in Computer Science*, pages 765–775. Springer Berlin Heidelberg, 2003.
22. L. Trevisan, G.B. Sorkin, M. Sudan, and D.P. Williamson. Gadgets, approximation, and linear programming. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 617–626, Oct 1996.
23. Y. Wang. Beyond baker's technique. *CoRR*, abs/1412.0340, 2014.
24. J. Yarkony. *Planarity Matters: MAP Inference in Planar Markov Random Fields with Applications to Computer Vision*. PhD thesis, University of California, Irvine, 2012.

A Elided Proofs

Theorem 1. *There is an $\alpha > 0$ such that, unless $P = NP$, there is no polynomial-time α-approximation algorithm for* Pairwise MAP MRF*, even for nonnegative ϕ and ψ.*

Proof. Maximum Cut is NP-hard to approximate to better than a 60/61 factor [22]. There is an approximation-preserving reduction from Max Cut to Pairwise MAP MRF, by setting $\phi_i(x_i) = 0$ and $\psi_{ij}(x_i, x_j)$ to be 1 if $x_i \neq x_j$ and 0 otherwise. ☐

Theorem 2. *The existence of an algorithm approximating* Pairwise MAP MRF *on planar graphs with maximum degree 4 and nonpositive ϕ_i and ψ_{ij} to any multiplicative factor implies $P = NP$.*

Proof. Proof of this theorem is a modification of a proof of a weaker theorem by Wang [23].

Given a planar graph G, we construct an Pairwise MAP MRF instance which has a score of 0 if and only if G is 3-colorable. As planar 3-colorability is NP-complete even on planar graphs of maximum degree 4 [9], and an approximation algorithm to a multiplicative factor must find a solution of weight 0 if one exists, this implies the theorem.

The Pairwise MAP MRF instance operates on G with $L = 3$ and functions

$$\begin{aligned} \phi_i(x) &= 0 && \forall\, x \in \{1,2,3\}, i \in V, \\ \psi_{i,j}(x,y) &= \begin{cases} 0 & \text{if } x \neq y \\ -1 & \text{if } x = y \end{cases} && \forall\, (i,j) \in E \end{aligned}$$

An assignment of score 0 is a 3-coloring where the labels are colors; the coloring is proper, as any edge with both endpoints of the same color would imply the value of the Pairwise MAP MRF instance is negative. Similarly, a 3-coloring induces an assignment of score 0. ☐

Corollary 1. *There is an EPTAS for* Correlation Clustering *in planar graphs.*

Proof. We present an approximation-preserving reduction from Correlation Clustering to Pairwise MAP MRF; with that, Theorem 3 gives the result.

Given an instance $\langle G, w, p\rangle$ of Correlation Clustering where G is planar, we construct an instance of Pairwise MAP MRF with the same graph, $L = 4$, $\phi_v(x_v) = 0$ for all $v \in V, x_v \in \{1,2,3,4\}$, and

$$\psi_{uv}(x_u, x_v) = \begin{cases} w(u,v) & \text{if } p(u,v) = 0 \text{ and } x_u = x_v \\ 0 & \text{if } p(u,v) = 0 \text{ and } x_u \neq x_v \\ 0 & \text{if } p(u,v) = 1 \text{ and } x_u = x_v \\ w(u,v) & \text{if } p(u,v) = 1 \text{ and } x_u \neq x_v \end{cases}.$$

If $\boldsymbol{x}$ is an assignment to this PAIRWISE MAP MRF instance, we make a cluster out of each maximal connected subgraph with the same label.

Edges with endpoints of different labels are exactly the edges between clusters, so the value of this CORRELATION CLUSTERING solution is the same as the value of $\boldsymbol{x}$.

In the other direction, we contract each cluster of a given partition down into a single supervertex to yield graph G', which is also planar. By the 4-color theorem, there exists an assignment of the labels $\{1, 2, 3, 4\}$ to the vertices of G' such that no adjacent vertices have the same label.

Give each vertex in G the same label as the corresponding supervertex in G'. Edges within a cluster have both edges corresponding to the same supervertex, and thus they have the same label. Edges between clusters have corresponding edges in G', and thus have endpoints with different labels. Thus the value of the assignment is exactly the value of the partition.

Both the creation of the corresponding PAIRWISE MAP MRF instance and the conversion of a solution of that instance to a solution of CORRELATION CLUSTERING take time linear in the size of the input. Thus there is a linear time approximation-preserving reduction, which, in conjunction with Theorem 3 completes the proof. Note that while the PTAS for PAIRWISE MAP MRF is *not* an *efficient* PTAS, this one is, because $L = 4 = O(1)$. □

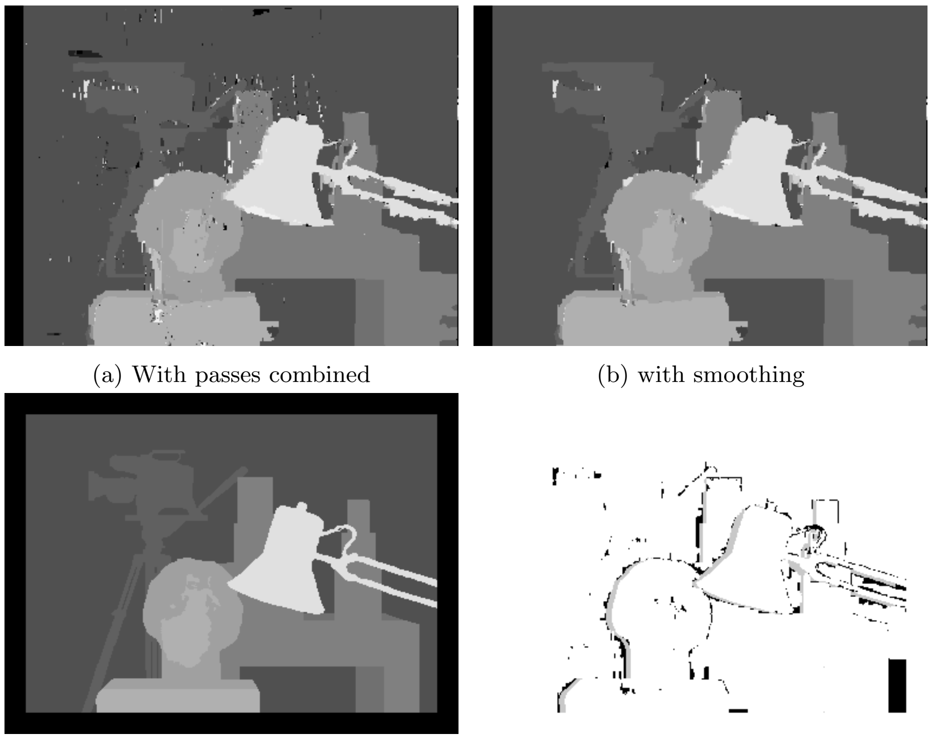
B Discussion of [23]

Lemma 4.2 of [23] is critical to the correctness of Wang's PTAS; as presented, it has some problems.

The stated runtime does not account for the degree of the graph; f_i has L^{d+1} possible outputs if vertex v_i has degree d; all possible outputs must be examined to ensure correctness.

Additionally, $S^U_{i \setminus p_i}$ is defined to be the max-sum of the liberal functions attached to vertices of $(U \cap V_{T_i}) \setminus (X_{p_i} \cup \delta X_{p_i})$. In a nice tree decomposition of a star, that resulting set is empty for all i except the root r, which means that the entire value of $S^U_{r \setminus p_r}$ is $\Gamma^{\sigma_{i \setminus p_i}}_{X_r \setminus X_{p_r}}$. $\Gamma^{\sigma_{i \setminus p_i}}_{X_r \setminus X_{p_r}}$, in this case, is defined to be the sum of liberal functions attached to every vertex in the star when the configuration of *just* the root is fixed to be $\sigma_{i \setminus p_i}$. So, calculating $\Gamma^{\sigma_{i \setminus p_i}}_{X_r \setminus X_{p_r}}$ is equivalent to solving the original problem and how it is calculated is not specified.

C Additional Figures



(a) With passes combined

(b) with smoothing

(c) Ground truth for comparison

(d) Mislabeled pixels highlighted

Fig. 4: Our results on *tsukuba* with heuristics applied.